Source Coding with Side Information at the Decoder: Models with Uncertainty, Performance Bounds, and Practical Coding Schemes

Elsa Dupraz*, Aline Roumy†, and Michel Kieffer*‡§
*LSS – CNRS – SUPELEC – Univ Paris-Sud, 91192 Gif-sur-Yvette, France, {firstname.name}@lss.supelec.fr
†INRIA, Campus de Beaulieu, 35042 Rennes, France, {firstname.name}@inria.fr
‡on leave at LTCI – CNRS - Telecom ParisTech, 75013 Paris
§Institut Universitaire de France

***Abstract*—We consider the problem of source coding with side information (SI) at the decoder only, when the joint distribution between the source and the SI is not perfectly known. Four parametric models for this joint distribution are considered, where uncertainty about the distribution is turned into uncertainty about the value of the parameters. More precisely, a prior distribution for the parameters may or may not be available. Moreover, the value of the parameters may either change at every symbol or remain constant for a while. This paper overviews the results on the performance of lossless source coding with SI at the decoder for three models, and gives the performance for the fourth. The way LDPC-based encoding and decoding schemes should be designed to cope with model uncertainty is provided. Most of the proposed practical schemes perform close to the theoretical limits.**

I. Introduction

Classical results on source coding with Side Information (SI) at the decoder only [17] as well as design of practical coding schemes for this problem [3], [13] rely on the assumption that the joint probability distribution between the source X and the SI Y is perfectly known. Nevertheless, this assumption is seldom satisfied in practical situations. It makes both the evaluation of the required coding rate and the design of a robust decoder quite difficult. This problem is usually mitigated via a feedback channel [1] or by allowing interactions between the encoder and the decoder [19]. Although such approaches may theoretically decrease the total coding rate, exchanges between the encoder and encoder lead to increased delay which may not be compatible with delay constrained applications. Consequently there is a need to characterize the performance of coding schemes with uncertainty in the joint distribution of (X, Y) in the case of one-way communication between the encoder and the decoder.

A universal coding setup is introduced in [8], where the distribution $P(X)$ of the source X is assumed unknown but the conditional distribution $P(Y|X)$ is perfectly known. Nevertheless, in many scenarios such as distributed video coding or distributed compression in network of sensors, $P(X)$ can be inferred at the encoder. However $P(Y|X)$ remains difficult to obtain accurately in a one-way communication. Similarly, in source coding with multiple SI [16], the joint distribution $P(X, Y)$ is uncertain at the encoder but available at the decoder. This assumption is also difficult to satisfy in practice, as obviously the decoder does not observe the source directly.

A way to capture the uncertainty is to consider source models in which the joint distribution $P(X, Y)$ belongs to a parametric family of joint distributions. Four models are considered in this paper. They represent different levels of knowledge and capture different types of time variations of the joint distribution $P(X, Y)$. For two models, the distribution of the source sequence $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ is parametrized by some unknown vector θ that is fixed for the sequence and that can vary from sequence to sequence, as in universal coding problems [4]. In the two other models, the distribution of (X_n, Y_n) is described by a vector π_n allowing model variations with n [2]. The distinction between θ fixed for a sequence, and the π_ns varying from symbol to symbol has already been proposed in the context of channel coding [12]. In both cases, a first model assumes that θ or the π_ns are realizations of some random variable Θ or of independent and identically distributed (i.i.d.) replicas of some random variable Π. A degraded model considers that no prior knowledge on θ or on the π_ns except their support is available. The latter choice is reasonable in all cases where a prior distribution for the parameters is difficult to obtain. For each model, the theoretical performance of the coding scheme may be given by a worst case. The question that arises is to determine the set in which this worst-case has to be searched for. The design of practical encoding and decoding schemes that would achieve the performance defined for these worst cases is also of interest, as the true values of the parameters are unknown.

This paper overviews the theoretical performance of lossless source coding with SI for three models and provides the performance limit for the fourth. An analysis of the model in which θ is described by a random variable is provided in [19] considering a variable-length setup. To gain a complete understanding of the problem, the fixed-length setup is considered here. It falls in the context of general sources [6] for which no closed form expression of the performance is available. Practical coding schemes for a Binary Symmetric Source (BSS) X and $P(Y|X)$ described by a Binary Symmetric Channel (BSC) are also introduced. The schemes are based on Low-Density Parity-Check (LDPC) codes, and account for the uncertainty in estimating the transition probability. Implementation issues are also discussed. More precisely, we define the rate at which the encoder should encode the source for a given outage constraint, we provide insights in the code design (i.e. code optimization), and finally the choice and the initialization of the decoder. The proposed schemes are shown to perform close to the theoretical limits: the best case is only 0.01 bit/symbol away from the theoretical rate. Related works are [4] which shows the existence of a universal linear code for lossless source coding with SI, and [11] which obtains a similar result for LDPC codes. Nevertheless none of these works give insights into the design of coding and decoding schemes that would achieve the optimal compression rate.

The paper is organized as follows. Definitions of the correlation models are given in Section II. Sections III and IV analyze the models with constant and time-varying parameter, respectively.

II. Signal model

In this paper, the source X to be compressed and the SI Y available at the decoder produce sequences of symbols $\{X_n\}_{n=1}^{+\infty}$ and $\{Y_n\}_{n=1}^{+\infty}$, respectively. $\mathcal{X}$ and $\mathcal{Y}$ denote the source and SI alphabets. Bold upper case letters, e.g. $\mathbf{X}_1^N = \{X_n\}_{n=1}^N$, denote random vectors, whereas bold lower case letters, $\mathbf{x}_1^N = \{x_n\}_{n=1}^N$, represent their realizations. Moreover, when it is clear from the context that the distribution of a random variable X_n does not depend on n, the index n is omitted.

The goal of this section is to model source uncertainty. Each of the four proposed models consists of a family of parametric distributions. Consider first the case of a time invariant parameter.

Definition 1. *(*P-Source*) A source (X,Y) with Prior (P-Source) produces a sequence of **independent** symbols $\{(X_n,Y_n)\}_{n=1}^{+\infty}$ drawn from a distribution belonging to a family $\{P(X,Y|\Theta=\theta)\}_{\theta\in\mathcal{P}_\theta}$ parametrized by a **random vector** Θ. The random vector Θ, with distribution $P_\Theta(\theta)$, takes its value in a set $\mathcal{P}_\theta$ that is either **discrete or continuous**. The source symbols X and Y take their values in **discrete sets** $\mathcal{X}$ and $\mathcal{Y}$ respectively. Moreover, the **realization of the parameter** θ **is fixed** for the sequence $\{(X_n,Y_n)\}_{n=1}^{+\infty}$.*

The P-source, determined by $\mathcal{P}_\theta$, $P_\Theta(\theta)$, and $\{P(X,Y|\Theta=\theta)\}_{\theta\in\mathcal{P}_\theta}$, is stationary but non-ergodic [5, Section 3.5].

Definition 2. *(*WP-Source*) A source (X,Y) Without Prior (WP-Source) produces a sequence of **independent** symbols $\{(X_n,Y_n)\}_{n=1}^{+\infty}$ drawn from a distribution belonging to a family $\{P_\theta(X,Y)\}_{\theta\in\mathcal{P}_\theta}$ parametrized by a vector θ. The vector θ takes its value in a set $\mathcal{P}_\theta$ that is either **discrete or continuous**. The source symbols X and Y take their values in **discrete sets** $\mathcal{X}$ and $\mathcal{Y}$ respectively. Moreover, the **parameter** θ **is fixed** for the sequence $\{(X_n,Y_n)\}_{n=1}^{+\infty}$.*

The WP-source, completely determined by $\mathcal{P}_\theta$ and $\{P_\theta(X,Y)\}_{\theta\in\mathcal{P}_\theta}$, is stationary but non-ergodic [5, Section 3.5]. The only difference between the P- and WP-Sources lies in the definition of θ. In the WP-Source, no distribution for θ is specified, either because its distribution is not known or because θ is not modeled as a random variable. The two next models allow parameter variations from symbol to symbol.

Definition 3. *(*M-Source*) A Mixture Source (X,Y), or M-Source, produces a sequence of **independent** symbols $\{(X_n,Y_n)\}_{n=1}^{+\infty}$ drawn $\forall n$ from $P(X_n,Y_n)$ that belongs to a family of distributions $\{P(X,Y|\Pi=\pi)\}_{\pi\in\mathcal{P}_\pi}$ **parametrized by a random vector** Π_n. The $\{\Pi_n\}_{n=1}^{+\infty}$ are i.i.d. with distribution $P(\Pi)$ and take their values in a **discrete set** $\mathcal{P}_\pi$. The source symbols X_n and Y_n take their values in **discrete sets** $\mathcal{X}$ and $\mathcal{Y}$, respectively.*

The M-Source, completely determined by $\mathcal{P}_\pi$, $P(\Pi)$, and $\{P(X,Y|\Pi=\pi)\}_{\pi\in\mathcal{P}_\pi}$, is stationary and ergodic [5, Section 3.5].

Definition 4. *(*WPM-Source*) A Without Prior Mixture Source (X,Y), or WPM-Source, produces a sequence of **independent** symbols $\{(X_n,Y_n)\}_{n=1}^{+\infty}$ drawn $\forall n$ from $P(X_n,Y_n)$ that belongs to a family of distributions $\{P_\pi(X,Y)\}_{\pi\in\mathcal{P}_\pi}$ **parametrized by a vector** π_n. The vectors π_n take their values in a **discrete set** $\mathcal{P}_\pi$. The source symbols X and Y take their values in **discrete sets** $\mathcal{X}$ and $\mathcal{Y}$, respectively.*

The WPM-Source, determined by $\mathcal{P}_\pi$ and $\{P_\pi(X,Y)\}_{\pi\in\mathcal{P}_\pi}$, is non-stationary and non-ergodic [5]. The only difference between the M and WPM-Source lies in the definition of the parameters π_n (no distribution for the π_n's is specified in the WPM-Model).

III. Time invariant parameters

This section focuses on the P and WP-Sources.

A. P-Source

Let us first define the coding functions for the P-Source.

Definition 5. *Let (X,Y) be a P-Source. Let $\mathcal{M}_N=\{1\dots|\mathcal{M}_N|\}$ be a set of integers. A coding process is defined by an encoding mapping $\phi_N:\mathcal{X}^N\to\mathcal{M}_N$ and a decoding mapping $\psi_N:\mathcal{M}_N\times\mathcal{Y}^N\to\mathcal{X}^N$. The error probability is*

$$P_e^N = P\left(\mathbf{X}_1^N \neq \psi_N(\phi_N(\mathbf{X}_1^N),\mathbf{Y}_1^N)\right) \tag{1}$$

A rate R is said to be achievable if and only if there exists a (ϕ_N,ψ_N)-code such that $\lim_{N\to\infty}P_e^N = 0$ and $\limsup_{N\to\infty}\frac{1}{N}\log|\mathcal{M}_N|\leqslant R$.

The following lemma gives the infimum of achievable rates for the P-Source.

Lemma 1. *The infimum of achievable rates $R_{X|Y}^P$ for the P-Source (X,Y) is given by (see Definition 5)*

$$R_{X|Y}^P = \sup_{\theta\in\mathcal{P}_\theta} H(X|Y,\Theta=\theta) \tag{2}$$

where $H(X|Y,\Theta=\theta)$ is obtained from

$$P(X|Y,\Theta=\theta) = \frac{P(X,Y|\Theta=\theta)}{\sum_{x'\in\mathcal{X}}P(X=x',Y|\Theta=\theta)}\ . \tag{3}$$

Proof: First (X,Y) is a general source because the parameter Θ is a random variable and the statistics of the random variables $P(\mathbf{X}_1^N,\mathbf{Y}_1^N)$ are well defined, $\forall N$. Therefore, the infimum of achievable rates can be derived with information spectrum tools [6, Section 7.2]:

$$\begin{aligned}R_{X|Y}^P &= \mathrm{p}-\limsup_{N\to\infty}\frac{1}{N}\log\frac{1}{P(\mathbf{X}_1^N|\mathbf{Y}_1^N)}\\ &:= \inf\left\{\alpha|\lim_{N\to\infty}P\left(\frac{1}{N}\log\frac{1}{P(\mathbf{X}_1^N|\mathbf{Y}_1^N)}>\alpha\right)=0\right\}\end{aligned}$$

However, for a fixed θ, the source is ergodic and

$$\frac{1}{N}\log\frac{1}{P(\mathbf{X}_1^N|\mathbf{Y}_1^N,\Theta=\theta)}\xrightarrow[N\to\infty]{\text{in proba.}}H(X|Y,\Theta=\theta).$$

Therefore,

$$\begin{aligned}R_{X|Y}^P &= \inf\left\{\alpha|\forall\theta\in\mathcal{P}_\theta,\lim_{N\to\infty}P(\tfrac{1}{N}\log\tfrac{1}{P(\mathbf{X}_1^N|\mathbf{Y}_1^N,\Theta=\theta)}>\alpha)=0\right\}\\ &= \inf\left\{\alpha|\forall\theta\in\mathcal{P}_\theta,\alpha\geqslant H(X|Y,\Theta=\theta)\right\}\\ &= \inf\left\{\alpha|\alpha\geqslant\sup_{\theta\in\mathcal{P}_\theta}(H(X|Y,\Theta=\theta))\right\}\\ &= \sup_{\theta\in\mathcal{P}_\theta}H(X|Y,\Theta=\theta).\end{aligned}$$

■

Corollary 1. *If the set $\mathcal{P}_\theta$ contains some θ such that X and Y are independent, then SI at the decoder does not reduce the infimum of achievable rates ($R_{X|Y}^P=H(X)$).*

To analyze the performance of the coding system we propose below, we also give the performance of the source coding scheme when an estimate $\hat\theta$ of θ is available at the decoder.

Lemma 2. *Let (X,Y) be a P-source and let $\hat\theta$ be a noisy version of θ available at the decoder and obtained from the known conditional distribution $P_{\hat\Theta|\Theta=\theta}(\hat\theta)$. Denote by $\mathcal{P}_{\hat\theta}\subseteq\mathcal{P}_\theta$ the set of every possible $\hat\theta$. The coding scheme is defined as in Definition 5, except for the*

mapping ψ_N that becomes $\psi_N : \mathcal{M}_N \times \mathcal{Y}^N \times \mathcal{P}_{\hat{\theta}} \to \mathcal{X}^N$. The infimum of achievable rates $R^P_{X|Y,\hat{\Theta}}$ in this setup is given by

$$R^P_{X|Y,\hat{\Theta}} = \sup_{\theta \in \mathcal{P}_\theta} H(X|Y, \Theta = \theta) \ . \tag{4}$$

Proof: The coding scheme can at least achieve the coding performance of a scheme without parameter estimate $\hat{\theta}$ (derived in Lemma 1). Therefore, $R^P_{X|Y,\hat{\Theta}} \leqslant \sup_{\theta \in \mathcal{P}_\theta} H(X|Y, \Theta = \theta)$.

Moreover, if the encoder transmits a source sequence with rate $R < \sup_{\theta \in \mathcal{P}_\theta} H(X|Y, \Theta = \theta)$, then the decoder fails to decode all sequences with parameter $\theta^\star$ such that $H(X|Y, \theta^\star) > R$ and $R^P_{X|Y,\hat{\Theta}} \geqslant \sup_{\theta \in \mathcal{P}_\theta} H(X|Y, \Theta = \theta)$. ∎

Lemma 2 shows that the knowledge of an estimate $\hat{\theta}$ of θ at the decoder does not decrease the coding rate. However, it enables to implement practical decoding schemes in a more convenient way, as described below.

We now provide practical coding solutions for the P-Source, based on LDPC codes [15]. Let X be a BSS and let Y be the output of a BSC, with $P(Y = 1|X = 0) = \theta$ and $\theta \in \mathcal{P}_\theta$. This channel is denoted BSC(θ). Let us first review how LDPC codes [10] are used to solve the source coding problem with SI at the decoder, when all source distributions are known at both encoder and decoder. For a given source sequence $\mathbf{x}_1^N$, a codeword $\mathbf{z}_1^M$ of length M is built as $\mathbf{z}_1^M = A^T \mathbf{x}_1^N$ where $A \in \mathcal{X}^{N \times M}$ is a binary sparse matrix. At the decoder, a dependency graph between the entries of $\mathbf{x}_1^N$, $\mathbf{y}_1^N$, and $\mathbf{z}_1^M$ is built using A. The graph is defined by check node and variable node degree distributions, respectively $\rho(x)$ and $\lambda(x)$, which satisfy the code rate constraint $M/N = \int_0^1 \rho(x)dx / \int_0^1 \lambda(x)dx$. An LDPC code with distributions $\rho(x)$ and $\lambda(x)$ is called a (λ, ρ)-code. The Maximum *a Posteriori* (MAP) estimate

$$\hat{x}_n = \arg\max_{x \in \mathcal{X}} P(X_n = x | \mathbf{Y}_1^N = \mathbf{y}_1^N) \tag{5}$$

is well approximated by a Message Passing (MP) algorithm performed in the graph. Various approximate MAP estimators are proposed in [14]. The soft LDPC decoder is a MP taking the conditional distributions $P(x_n|y_n)$ as input messages while hard LDPC decoders only require the value of the SI sequence $\mathbf{y}_1^N$ as input, at the price of lower performance.

With the P-Source, as we do not know precisely the true conditional distribution $P(X|Y)$, we propose a two-stage coding/decoding setup. First, a subsequence $\mathbf{x}_1^{u_N}$ of $\mathbf{x}_1^N$ is LDPC coded. Since the statistics are not yet known, this learning sequence is decoded with a hard LDPC decoder[1]. In the binary case, the rate R_l at which the learning sequence has to be encoded, can be evaluated with density evolution [14] and is dimensioned for the worst $\theta \in \mathcal{P}_\theta$. The Bayesian estimation of θ from $(\mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N})$ provides the posterior distribution $P_{\Theta|\mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N}}(\theta)$ [7].

Second, for the source symbols X_{u_n} to X_N, (5) becomes

$$\hat{x}_n = \arg\max_{x \in \mathcal{X}} P(X_n = x | \mathbf{Y}_{u_N+1}^N = \mathbf{y}_{u_N+1}^N, \mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N}) \tag{6}$$

An approximation of (6) is obtained by the soft LDPC decoder now initialized by

$$P(X_n|Y_n, \mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N}) = \int_{\theta \in \mathcal{P}_\theta} P(X_n|Y_n, \theta) P_{\Theta|\mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N}}(\theta) d\theta. \tag{7}$$

In the binary case, the density (7) is completely determined by a reconstruction parameter θ_r with $\theta_r = P(X_n = 1|Y_n = 0, \mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N})$.[2] Note that in place of (7), one could initialize the decoder with $P(X_n|Y_n = y_n, \hat{\theta})$, where $\hat{\theta}$ is the MAP estimate of θ. However this does not take into account the uncertainty on $\hat{\theta}$. Instead, θ_r and therefore the MP algorithm account for the uncertainty in estimating θ since (7) contains an integration with posterior distribution $P_{\Theta|\mathbf{x}_1^{u_N}, \mathbf{y}_1^{u_N}}(\theta)$.

To evaluate the performance of the second part of the coding scheme, density evolution [14] is used to provide for a given (λ, ρ)-code, the largest $\theta^\star$ s.t. $\mathbf{x}_1^N$ can be decoded with probability of error less than a target value ε_t. In channel coding, density evolution holds for symmetric distributions but in Slepian Wolf coding, it can be generalized to non symmetric binary sources [3]. $\theta^\star$ is called the threshold of the code. To obtain $\theta^\star$, the only information needed by the density evolution algorithm is the probability density function of the input messages of the decoder. For a BSC(θ), it is [14]

$$p_X(x) = \theta\delta\left(x + \log\frac{1-\theta}{\theta}\right) + (1-\theta)\delta\left(x - \log\frac{1-\theta}{\theta}\right) \tag{8}$$

where δ refers to the Dirac distribution.

As $\theta \neq \theta_r$, the input messages (7) are not correct and our scheme is a mismatch decoder. To take this mismatch into account, density evolution has to be initialized with

$$p_X(x) = \theta\delta\left(x + \log\frac{1-\theta_r}{\theta_r}\right) + (1-\theta)\delta\left(x - \log\frac{1-\theta_r}{\theta_r}\right). \tag{9}$$

Actually, the log-likelihood ratio is initialized with $\log\frac{1-\theta_r}{\theta_r}$ (if $Y = 1$) and with its opposite (if $Y = 0$), and this occurs with probability θ and $1-\theta$, respectively.

To complete our scheme definition, we now compute the coding rate for the remaining $N - u_N$ symbols. Ensuring error probability less than ε_t for every $\theta \in \mathcal{P}_\theta$ would lead to an important rate increase. In what follows, an outage parameter γ is introduced. The outage authorizes the decoder to fail for a subset of $\mathcal{P}_\theta$ of measure γ.

Definition 6. *For a given $\gamma \in [0, 1]$ and θ_r, consider a set $B^\gamma_{\theta|\theta_r} \subseteq \mathcal{P}_\theta$ such that*

$$\int_{\theta \in \mathcal{B}^\gamma_{\theta|\theta_r}} P_{\Theta|\theta_r}(\theta) d\theta > 1 - \gamma \ . \tag{10}$$

1) *For a (λ, ρ)-code and a given ε_t, a **mismatch rate** $R^{B^\gamma_{\theta|\theta_r}}_{\lambda,\rho,\varepsilon_t}$ is such that every source with parameter $\theta \in B^\gamma_{\theta|\theta_r}$ can be decoded using a reconstruction parameter θ_r and with an error probability less than ε_t.*
2) *A **reconstruction rate** $R^\gamma_{\lambda,\rho,\varepsilon_t}(\theta_r)$ is such that there exists a set $B^\gamma_{\theta|\theta_r} \subseteq \mathcal{P}_\theta$ for which $R^{B^\gamma_{\theta|\theta_r}}_{\lambda,\rho,\varepsilon_t}$ is a mismatch rate for $B^\gamma_{\theta|\theta_r}$.*

From Definition 6, the rate needed to transmit the source with target error probability ε_t and outage γ is

$$R^N_c(\gamma, \varepsilon_t) = \sup_{\theta_r \in \mathrm{Conv}(\mathcal{P}_\theta)} \inf_{B^\gamma_{\theta|\theta_r}} \inf_{\lambda,\rho} R^\gamma_{\lambda,\rho,\varepsilon_t}(\theta_r) \tag{11}$$

In fact, the inf on (λ, ρ) corresponds to the design of a good LDPC code. Then, for a given γ, there are many subsets $B^\gamma_{\theta|\theta_r}$ of $\mathcal{P}_\theta$ of measure γ. The outage condition allows the decoder to fail for one of the subsets. Therefore we seek for the most advantageous subset, i.e. the subset that minimizes the coding rate. This gives the inf on $B^\gamma_{\theta|\theta_r}$. Finally, as θ_r is not known at the encoder, the rate has to be dimensioned for the worst possible case, which gives the sup on θ_r.

[1] Note that a learning sequence in this setup differs from a learning sequence in channel coding since this sequence contains useful data.

[2] θ_r does not necessarily belong to $\mathcal{P}_\theta$. For a BSC(θ), $\theta_r \in \mathrm{Conv}(\mathcal{P}_\theta)$, the convex hull of the elements of $\mathcal{P}_\theta$.

TABLE I
THEORETICAL (TH.) AND PRACTICAL (PRAC.) RATE BOUNDS IN BIT/SYMBOL WHEN X IS A BSS AND $P(Y|X)$ A BSC.

Source	Conditions	Th. Rate	Prac. Rate
P-Source	$\mathcal{P}_\theta = [0.1, 0.215]$	0.74	0.75
WP-Source	$\mathcal{P}_\theta = [0.1, 0.215]$	0.74	0.75
M-Source	$\mathcal{P}_\pi = \{0.1, 0.215\}$ $p = 0.143$	0.59	0.6
WPM-Source	$\mathcal{P}_\pi = \{0.1, 0.143\}$	0.59	0.75
No SI	$P(X=1) = 0.5$	1	1

The rate of the whole coding system for fixed γ and ε_t is

$$R^N(\gamma, \varepsilon_t) = \frac{u_N}{N} R_l + \frac{N - u_N}{N} R_c^N(\gamma, \varepsilon_t) \ . \tag{12}$$

Although difficult to solve in practice, the optimization problem (11) expresses the tradeoff between the length of the learning sequence, the outage parameter, and the rate performance thus giving insights to design a practical coding solution.

For $\lim_{N\to\infty} u_N = +\infty$, the Bayesian estimator is consistent [9, Section 11.6], $\hat{\Theta}$ converges in probability to the true θ. Hence the outage condition is no more useful, as every θ is estimated perfectly. By setting $\lim_{N\to\infty} u_N = +\infty$ while $\lim_{N\to\infty} u_N/N = 0$, (12) becomes the rate needed to transmit a source with the worst parameter θ in (4). This asymptotic rate depends only of the chosen (λ, ρ)-code and can be very close to the entropy [15].

Example 1. *Let X be a BSS and let Y be the output of a BSC(θ) with input X. $\mathcal{P}_\theta = [0.1, 0.215]$ and $P_\Theta(\theta)$ is piece-wise constant on $\mathcal{P}_\theta$. For instance, $P_\Theta(\theta) = 10\ \mathbb{I}_{\theta\in[0.1,0.15]} + 8\ \mathbb{I}_{\theta\in[0.15,0.20]} + 6.67\ \mathbb{I}_{\theta\in[0.20,0.215]}$ where $\mathbb{I}$ is the indicator function. From Lemma 1, the infimum achievable rate is $R^P_{X|Y} = 0.74$ bit/symbol. With an LDPC code of distributions $\rho(x) = x^5$ and $\lambda(x) = 0.093x^3 + 0.720x^4 + 0.187x^5$ obtained with a differential evolution algorithm [18] for the worst parameter $\theta \in \mathcal{P}_\theta$, a rate $R = 0.75$ bit/symbol is achieved for $\varepsilon_t = 10^{-5}$, see Table III-A. Without SI available at the decoder, the infimum of achievable rates is 1 bit/symbol, since $P(X=1) = 0.5$.*

B. WP-Sources

Now, since only the support for θ is known, information spectrum approaches do not apply.

Definition 7. *Let (X, Y) be a WP-Source. The mappings ϕ_N and ψ_N are as in Definition 5. The probability of error for a given θ is*

$$P_e^N(\theta) = P_\theta\left(\mathbf{X}_1^N \neq \psi_N(\phi_N(\mathbf{X}_1^N), \mathbf{Y}_1^N)\right) \ . \tag{13}$$

A rate R is said to be achievable if and only if there exists a (ϕ_N, ψ_N) code such that $\forall\theta \in \mathcal{P}_\theta$, $\lim_{N\to\infty} P_e^N(\theta) = 0$ and $\limsup_{N\to\infty} \frac{1}{N} \log|\mathcal{M}_N| \leqslant R$.

Lemma 3 (see [4])**.** *For the WP-Source, the infimum of achievable rates is*

$$R^{WP}_{X|Y} = \sup_{\theta\in\mathcal{P}_\theta} H_\theta(X|Y) \tag{14}$$

where $H_\theta(X|Y)$ is calculated from

$$P_\theta(X|Y) = \frac{P_\theta(X,Y)}{\sum_{x'\in\mathcal{X}} P_\theta(X=x', Y)} \ . \tag{15}$$

Corollary 2. *The knowledge of a prior for θ does not reduce the infimum of achievable rates (if θ has same support for both WP and P sources, $R^{WP}_{X|Y} = R^P_{X|Y}$).*

An intuitive argument is that a sequence with parameter θ needs a rate $H_\theta(X|Y)$ to be correctly decoded. Since the encoder has no way to predict the exact value of the parameter θ of a sequence, for the P-Source, as well as for the WP-Source, it is forced to protect the sequence against the worst parameter and to dimension the rate consequently.

The practical coding solution proposed here is adapted from that of Section III-A. The main difference is on the evaluation of $\hat{\theta}$. Indeed, here, the variable θ is deterministic and hence is estimated by performing Maximum Likelihood (ML) estimation on the learning sequence. In the second phase, the distribution provided to the decoder is directly $P_{\hat{\theta}}(X|Y)$. Its asymptotic conditional distribution $P_\theta(\hat{\Theta})$ for a given θ is taken from [7, Section 8.2.2]

$$\hat{\Theta}|\theta \sim \mathcal{N}(\theta, i(\theta)) \tag{16}$$

where $i(\theta)$ is the Fisher information. From [7, Section 8.2.2], (16) can be approximated as $\mathcal{N}(\hat{\theta}, i(\hat{\theta}))$. The outage condition is now defined as follows.

Definition 8. *For a given $\gamma \in [0,1]$, let $\alpha^{(1-\gamma)}$ be the $(1-\gamma)$-th percentile of a Gaussian distribution. Consider an estimate $\hat{\theta}$ of θ and $B^\gamma_{\theta|\hat{\theta}} = \{\theta \in \mathcal{P}_\theta, |\theta - \hat{\theta}| < \alpha^{(1-\gamma)} i(\hat{\theta})\}$.*

1) *For a (λ, ρ)-code and a target error probability ε_t, a **mismatch rate** $R^{B^\gamma_{\theta|\hat{\theta}}}_{\lambda,\rho,\varepsilon_t}(\hat{\theta})$ is such that every source with parameter $\theta \in B^\gamma_{\theta|\hat{\theta}}$ can be decoded using $\hat{\theta}$ at the decoder with an error probability less than ε_t.*
2) *A **reconstruction rate** $R^\gamma_{\lambda,\rho,\varepsilon_t}(\hat{\theta})$ is such that there exists a set $B^\gamma_{\theta|\hat{\theta}} \subseteq \mathcal{P}_\theta$ such that $R^\gamma_{\lambda,\rho,\varepsilon_t}(\hat{\theta})$ is a mismatched rate for $B^\gamma_{\theta|\hat{\theta}}$.*

From Definition 8, the rate needed to transmit the source with outage parameter γ and target error probability ε_t is

$$R_c^N(\gamma, \varepsilon_t) = \sup_{\hat{\theta}\in\mathcal{P}_\theta} \inf_{B^\gamma_{\theta|\hat{\theta}}} \inf_{\lambda,\rho} R^\gamma_{\lambda,\rho,\varepsilon_t}(\hat{\theta}) \ . \tag{17}$$

To finish, R_l and $R^N(\gamma, \varepsilon_t)$ are defined as in Section III-A. As the ML estimator is consistent [9], asymptotic considerations on the rate hold also in this setup.

Example 2. *The model of Example 1 is considered, without prior on θ. The same rate $R = 0.74\ bit/symbol$ is asymptotically achieved with the same (λ, ρ)-code.*

IV. TIME-VARYING PARAMETERS

This section focuses on the M- and WPM-Sources.

A. M Source

For the M-Source, the distribution $P(\Pi)$ is perfectly known. The source symbols (X, Y) are i.i.d. and the joint distribution $P(X, Y)$ is perfectly determined as

$$P(X,Y) = \sum_{\pi\in\mathcal{P}_\pi} P(\pi) P(X,Y|\pi) \ . \tag{18}$$

The source is stationary and ergodic and the results on lossless source coding with SI introduced in [17] apply. Then, one has

Lemma 4. [17] *The infimum achievable rate for source coding with SI at the decoder with the M-Source is*

$$R^M_{X|Y} = H(X|Y) \tag{19}$$

where $H(X|Y)$ is calculated from (18).

Since the statistics of the source are perfectly known, soft LDPC decoding can be applied directly. The optimization and performance of the code is obtained via density evolution, using $P(X|Y)$.

Example 3. *Here again, X is a BSS. The correlation between X and Y is described by a BSC(π) with random transition probability. $\mathcal{P}_\pi = \{\pi_1 = 0.1, \pi_2 = 0.215\}$ with $P(\Pi = \pi_1) = 0.61$ and $P(\Pi = \pi_2) = 0.39$. This gives a conditional distribution $p = P(Y = 1|X = 0) = 0.143$. One obtains an achievable rate $R^M_{X|Y} = 0.59$ bit/symbol. From an optimization with differential evolution, the degree distributions $\rho(x) = x^5$ and $\lambda(x) = 0.099x^3 + 0.712x^4 + 0.174x^5 + 0.015x^6$ is obtained and gives an achievable rate $R = 0.60$ bit/symbol.*

B. Mixture Source Without Prior

Consider now the WPM-Source, in which no prior information except the support is known about the sequence of parameters $\{\pi_n\}_{n=1}^N$.

Definition 9. *Consider a WPM-Source and mappings ϕ_N and ψ_N introduced in Definition 5. The probability of error is defined for a given sequence $\{\pi_n\}_{n=1}^N \in \mathcal{P}_\pi^N$ as*

$$P_e^N(\{\pi_n\}_{n=1}^N) = P_{\{\pi\}_{n=1}^N}\left(\mathbf{X}_1^N \neq \psi_N(\phi_N(\mathbf{X}_1^N), \mathbf{Y}_1^N)\right) . \quad (20)$$

A rate R is said to be achievable if and only if there exists a (ϕ_N, ψ_N)-code such that $\forall\{\pi_n\}_{n=1}^N \in \mathcal{P}_\pi^N$,

$$\lim_{N\to\infty} P_e^N(\{\pi_n\}_{n=1}^N) = 0 \quad (21)$$

and $\limsup_{N\to\infty} \frac{1}{N} \log |\mathcal{M}_N| < R$.

Lemma 5. [2] *Let (X, Y) be a WPM-Source and let the coding scheme be as defined in Definition 9. The infimum achievable rate for source coding with SI at the decoder is*

$$R^{WPM}_{X|Y} = \sup_{P(X,Y)\in Conv\left(\{P_\pi(X,Y)\}_{\pi\in\mathcal{P}_\pi}\right)} H(X|Y) \quad (22)$$

where Conv$(\{P_\pi(X, Y)\}_{\pi\in\mathcal{P}_\pi})$ is the convex hull of the elements of $\{P_\pi(X, Y)\}_{\pi\in\mathcal{P}_\pi}$.

Corollary 3. *If the convex hull of the elements of the set $\mathcal{P}_\pi$ contains some π such that X and Y are independent, then SI at the decoder does not reduce the infimum of achievable rates ($R^P_{X|Y} = H(X)$).*

A coding scheme for the WPM-Source is by far the most complicated to realize in practice. Indeed, as the process is non-stationary, learning the statistics of the sources is useless, and hence it is not possible to perform soft LDPC decoding as in Section III-A. Nevertheless, hard LDPC decoding enables to code and decode such a source. For a given (λ, ρ)-code, an associated rate $R_{\lambda,\rho}$ and a target error probability ε_t, density evolution gives us a threshold $\pi^\star$ such that, $\forall\pi \leqslant \pi^\star$, the hard LDPC decoder can decode a sequence with true parameter π with error probability less than ε_t. The (λ, ρ)-code for our setup is hence chosen such that its threshold $\pi^\star$ is up to the worst possible distribution in (22)

Example 4. *The set $\mathcal{P}_\pi$ contains now two probability transitions $\pi_1 = 0.1$ and $\pi_2 = 0.143$. The sup in (22) is $\pi_2 = 0.143$, giving a minimum achievable rate $R^{TVM}_{X|Y} = 0.59$ bit/symbol, as in Example 3. But now, with the hard LDPC decoding algorithm E [15], the rate achieved in practice is $R = 0.75$ bit/symbol with a regular LDPC code $3/4$.*

V. Conclusion

This paper addresses the problem of lossless source coding with SI at the decoder when the joint distribution of the source and side information is only partially known. Four parametric models have been considered, leading to different performance limits and different coding strategies, even if all consider LDPC codes.

This paper also provides some insights on implementation issues for the uncertainty aware schemes. More precisely, the practical encoding rate has been defined, when outage is allowed. Moreover, LDPC codes have been designed, and the choice and the initialization of the decoder have been presented.

VI. Acknowledgments

The authors would like to thank Valentin Savin and Mael Letreust for their helpful comments and advice.

References

[1] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, volume 1, pages 240–244, 2002.

[2] R. Ahlswede. Coloring hypergraphs: A new approach to multi-user source coding-1. *Journal of Combinatorics*, 4(1):76–115, 1979.

[3] J. Chen, D.K. He, and A. Jagmohan. The equivalence between Slepian-Wolf coding and channel coding under density evolution. *IEEE Trans. on Comm.*, 57(9):2534–2540, 2009.

[4] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Trans. on Inf. Th.*, 28(4):585–592, 1982.

[5] R.G. Gallager. *Information theory and reliable communication*. Wiley, 1968.

[6] T.S. Han. *Information-spectrum methods in information theory*. Springer, 2003.

[7] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2009.

[8] S. Jalali, S. Verdú, and T. Weissman. A Universal Scheme for Wyner-Ziv Coding of Discrete Sources. *IEEE Trans. on Inf. Th.*, 56(4):1737–1750, 2010.

[9] S.M. Kay. *Fundamentals of Statistical Signal Processing, Estimation theory*. Prentice Hall PTR, 1993.

[10] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Comm. Letters*, 6:440–442, 2002.

[11] T. Matsuta, T. Uyematsu, and R. Matsumoto. Universal Slepian-Wolf source codes using Low-Density Parity-Check matrices. In *IEEE Int. Symp. on Inf. Th. Proceedings (ISIT), 2010*, pages 186 –190, june 2010.

[12] P. Piantanida, G. Matz, and P. Duhamel. Outage behavior of discrete memoryless channels under channel estimation errors. *IEEE Trans. on Inf. Th.*, 55(9):4221–4239, Sep 2009.

[13] R. Puri and K. Ramchandran. PRISM: A new robust video coding architecture based on distributed compression principles. In *Proc. of the Annual Allerton Conf. on Comm. Cont. and Comp.*, volume 40, pages 586–595. Citeseer, 2002.

[14] T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke. Design of capacity-approaching irregular Low-Density Parity-Check codes. *IEEE Trans. on Inf. Th.*, 47(2):619–637, 2001.

[15] T.J. Richardson and R.L. Urbanke. The capacity of Low-Density Parity-Check codes under message-passing decoding. *IEEE Transactions on Information Theory*, 47(2):599–618, 2001.

[16] A. Sgarro. Source coding with side information at several decoders. *IEEE Trans. on Inf. Th.*, 23(2):179–182, 1977.

[17] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. on Inf. Th.*, 19(4):471–480, July 1973.

[18] R. Storn and K. Price. Differential evolution– a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[19] E.H. Yang and D.K. He. Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder. *IEEE Trans. on Inf. Th.*, 56(4):1808–1824, 2010.